



2/6

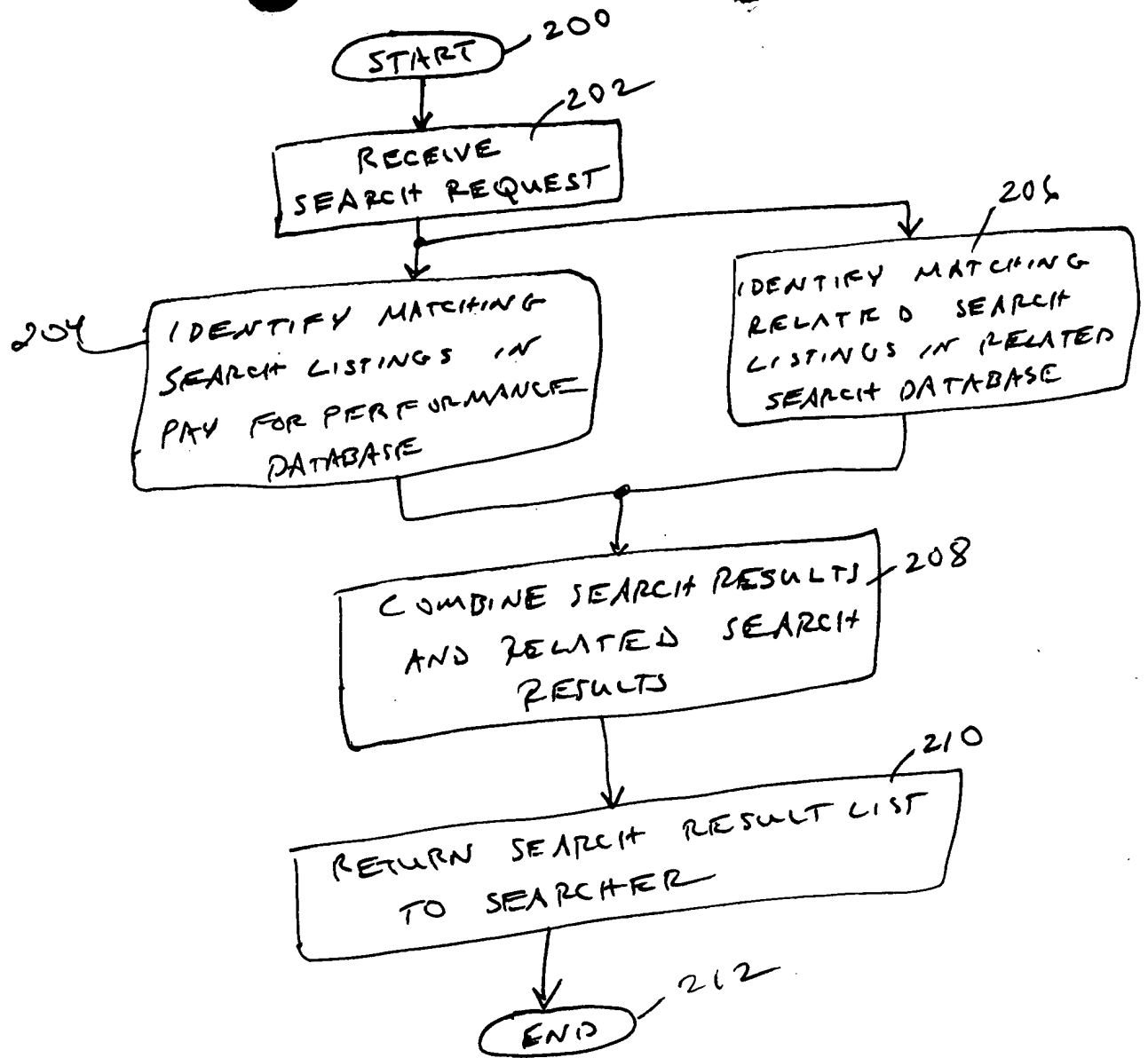


FIG. 2



[illegible]

FIG. 5

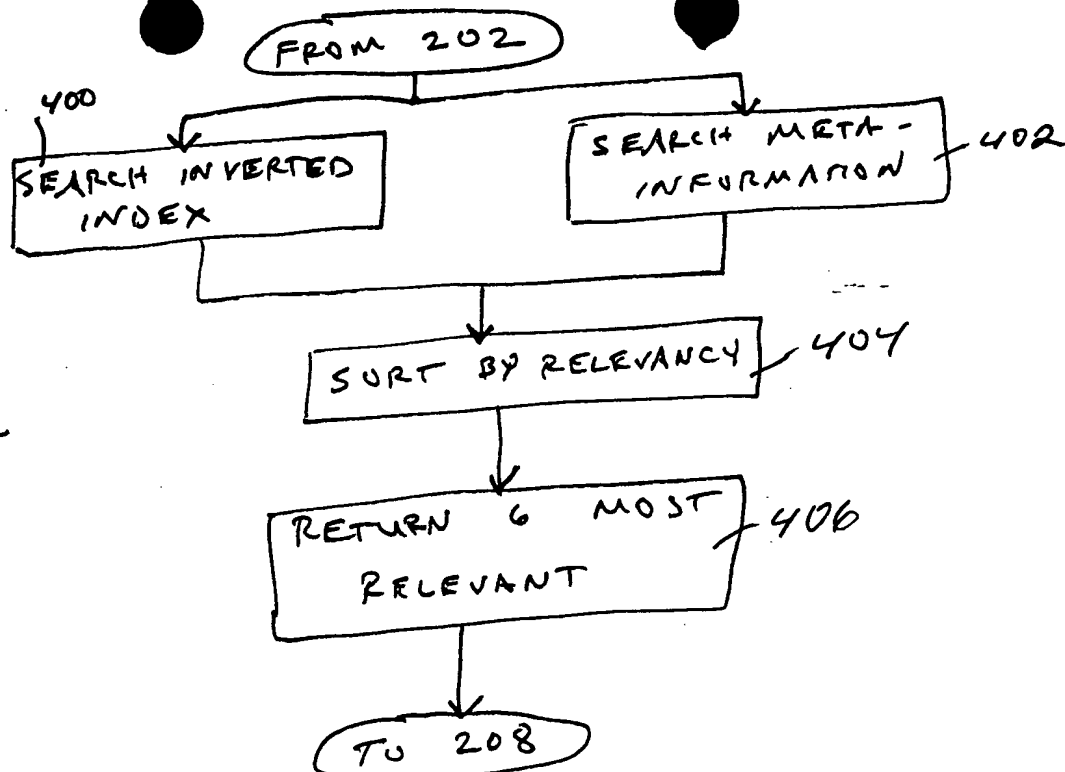
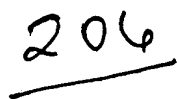


FIG. 4

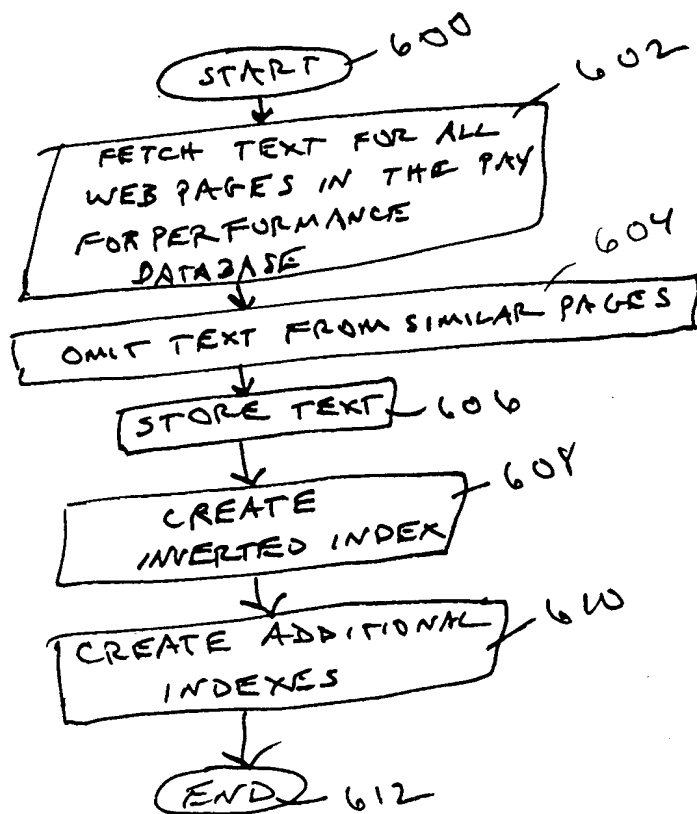


FIG. 6

6/6

FROM 602

Extract Complete URL list from Bidded Search Listing DB 702

Sort URL list and remove exact duplicates 704

Is this URL "similar" to the preceding URL in list? 706  
YES NO

Add to list of candidate duplicate URL's 708

Crawl 1st 2 of each potentially duplicate URL 710

Is URL body similar enough to previous URL body? 712  
YES NO

Don't crawl more of similar url's. Assign previously crawled body to rest of current set 714

Add to list to be crawled fully parallelized 718

Crawl all non-duplicate URL's 720

Load results into Free Text Database joined with search listing data 716

TO 606

604

FIG. 7